



## HSLs:09 Sample Design, Weights, Variance, and Missing Data

### Module Objectives

- Describe HSLs:09 weights that must be applied to assure estimates made from the data are representative of the study population
- Describe appropriate procedures for calculating standard errors

### HSLs:09 Sample Design

- HSLs:09 is a sample
  - Nationally representative of 9th graders in 2009-2010
  - Nationally representative of schools with 9th and 11th grades
- The sample for HSLs:09 is not a simple random sample (SRS) of the target population
  - HSLs:09 is a stratified, two-stage random sample design with primary sampling units (PSUs) defined as schools selected at the first stage and students randomly selected from schools at the second stage

### Sample Design: Two-stage Stratified Sample

- First stage = schools
  - Stratified random sample of public and private schools
    - ↓
    - 1,889 eligible
      - ↓
      - 944 (55.5% weighted) participated

### Sample Design: Two-stage Stratified Sample

- Second stage = 9th-grade students
  - Random sample from sampled schools' enrollment list
    - 25,206 eligible (about 27 per school)
      - Eligible students were capable of completing a student questionnaire and algebra assessment
      - Ineligible students (language barriers or severe disabilities) were retained in the sample and contextual data were obtained
      - 24,658 were classified as questionnaire-capable
      - 548 were classified as questionnaire-incapable
    - 21,444 (86% weighted) students participated
- Samples representative nationally and for 10 states

### Purpose of [Weights](#) – Review

- Weights are used to make estimates from the sample data representative of the target population
- Weights account for differential selection probabilities and differential patterns of response/nonresponse

**HSLs:09 Weights**

- Multiple weights are provided for analysis
- Weights account for nonresponse; ideally there would be a weight available that is adjusted for nonresponse to every component of every round of data collection
- Number of possible weights increases dramatically with longitudinal studies
- Researchers must decide which weight is the best for their research question

**HSLs:09 Nomenclature for Weight Variables**

- 1st Character = W, signifies "Weight variable"
- 2nd Character = 1, 2, etc., signifies a particular round of data collection

## HSLs:09 Weights

- The set of weights available does not include weights that account for nonresponse to every component or combination of components
- When no weight corresponds exactly to the combination of components included in the analysis, consider:
  - A weight with nonresponse adjustments for *more* components than are included in the analysis
    - May result in a slightly smaller analytic sample
    - Will adjust for nonresponse associated with each of the components that it covers
  - A weight with nonresponse adjustments for *fewer* components than are included in the analysis
    - May result in a larger analytic sample and bias

Variable Name	Nonresponse adjusted component(s) in each weight	HSLs:09 study round	Estimation
<b>School Analyses</b>			
W1SCHOOL	School	Base year	School-level, Base year only
<b>Student Analyses</b>			
W1STUDENT	Student	Base year	Base year only
W1PARENT	Student*Parent		
W1SCITCH	Student*Science Teacher		
W1MATHTCH	Student*Math Teacher		
W2STUDENT	Student	First follow-up	First follow-up only
W2PARENT	Parent		
W2W1STU	Student <sup>4</sup>	Base year and first follow-up	Change from base year to first follow-up
W2W1PAR	Student*Parent <sup>4</sup>		

<sup>4</sup> The student longitudinal weights account for nonresponse in the base year, the first follow-up, or both

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. High School Longitudinal Study of 2009 (HSLs: 09) First Follow-up Data File Documentation (NCES 2014-361).

## HSLs:09 Weights (Continued)

Variable Name	Nonresponse adjusted component(s) in each weight	HSLs:09 study round	Estimation
<b>School Analyses</b>			
W1SCHOOL	School	Base year	School level, Base year only
<b>Student Analyses</b>			
W1STUDENT	Student	Base year	Base year only
W1PARENT	Student*Parent		
W1SCITCH	Student*Science Teacher		
W1MATHTCH	Student*Math Teacher		
W2STUDENT	Student	First follow-up	First follow-up only
W2PARENT	Parent		
W2W1STU	Student <sup>4</sup>	Base year and first follow-up	Change from base year to first follow-up
W2W1PAR	Student*Parent <sup>4</sup>		

<sup>4</sup> The student longitudinal weights account for nonresponse in the base year, the first follow-up, or both

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. High School Longitudinal Study of 2009 (HSLs: 09) First Follow-up Data File Documentation (NCES 2014-361).

**Standard Error Calculation in HSLs:09 – Replication Techniques**

- This method calculates appropriate SEs based on differences between estimates from the full sample and a series of created subsamples (replicates)
- Select replicate weights that are associated with your main sampling weight (e.g., for W1STUDENT, select W1STUDENT001 through W1STUDENT200)
- HSLs:09 replication weights use the Balanced Repeated Replication (BRR) method

**Standard Error Calculation in HSLs:09 – Taylor Series Linearization**

- For HSLs:09, Taylor series linearization requires restricted-use data
- This method uses primary sampling unit (PSU) and strata identifiers to calculate appropriate SEs
- Select the PSU and stratum variables (variable names: PSU and STRAT\_ID) associated with your sampling weight variable

### Missing Data in HSLs:09

#### "Reserve" Codes

- -5 Data suppressed
- -7 Legitimate skip
- -8 Unit missing
- -9 Item missing

### Missing Data Example

- S1EDUEXPECT      S1 G01      How far in school 9th grader thinks he/she will get
- S1SURECLG      S1 G02      How sure 9th grader is that he/she will go to college to pursue a BA/BS

## Elementary and Secondary Education Data

S1EDUEXPECT	n	%	S1SURECLG AND S1EDUEXPECT	n	%	RECODE %
-9 Missing	308	1.4	-9 Missing	397	1.9	missing
1 less than high school	92	.4	-7 Item legitimate skip/NA	8546	3.9	missing
2 High school diploma or GED	2572	12.0	1 Very sure about going	9247	43.1	74.0
3 Start an Associate's degree	139	.6	2 Will probably go	3092	14.4	24.7
4 Complete an Associate's degree	1174	5.5	3 Will probably not go	129	.6	1.0
5 Start a Bachelor's degree	113	.5	4 Very sure about not going	33	.2	0.3
6 Complete a Bachelor's degree	3469	16.2	Total	21444	100.0	100.00
7 Start a Master's degree	226	1.1				
8 Complete a Master's degree	4214	19.7				
9 Start Ph.D./M.D./Law/other prof degree	172	.8				
10 Complete Ph.D./M.D./Law/other prof degree	4396	20.5				
11 Don't know	4569	21.3				
Total	21444	100.0				

  

S1SURECLG	n	%	S1SURECLG RELOADED	n	%	RECODE %
-9 Missing	397	1.9	-9 Missing	397	1.9	missing
-7 Item legitimate skip/NA	8546	39.9	(recode -7 to 0) Does not expect to go	8546	39.9	40.6
1 Very sure about going	9247	43.1	1 Very sure about going	9247	43.1	43.9
2 Will probably go	3092	14.4	2 Will probably go	3092	14.4	14.7
3 Will probably not go	129	.6	3 Will probably not go	129	.6	0.6
4 Very sure about not going	33	.2	4 Very sure about not going	33	.2	0.2
Total	21444	100.0	Total	21444	100.0	100.0

## Frequently Asked Questions

When selecting a weight, do I have to subset my dataset?

- No. The weight automatically limits your sample to cases with a positive weight

What happens to cases where there is no positive weight?

- They automatically drop out of your analytic sample

What weights do I use if I'm analyzing a subsample of cases?

- The same weights you would use when analyzing the full sample

What if I'm running a regression – what weights do I use?

- The same weights you would use for any other type of analysis



### Module Summary and Resources

#### Summary

- HSLs:09 weights that must be used to assure data are representative of the study population
- The procedures for calculating standard errors for HSLs:09 data

#### Resources

- [Analyzing NCES Complex Survey Data](#)
- [Statistical Analysis of NCES Datasets Employing a Complex Sample Design](#)